

TDDE13 LE3 HT2025

- Agentic AI

Fredrik Heintz

**Dept. of Computer Science
Linköping University**

fredrik.heintz@liu.se

@FredrikHeintz

Outline:

- **Large Language Models**
- **Agentic AI**
- **Model Context Protocol**

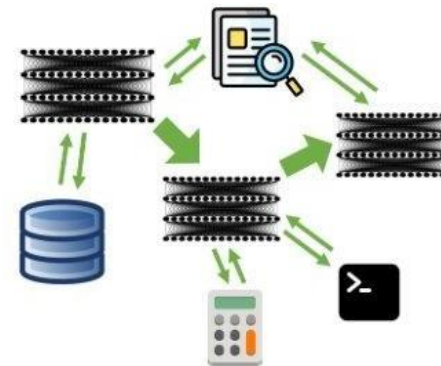
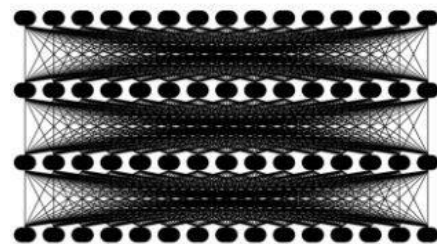
From Generative to Agentic

- Generative

- Generate content like text & image

- Agentic

- Execute complex tasks on behalf of human



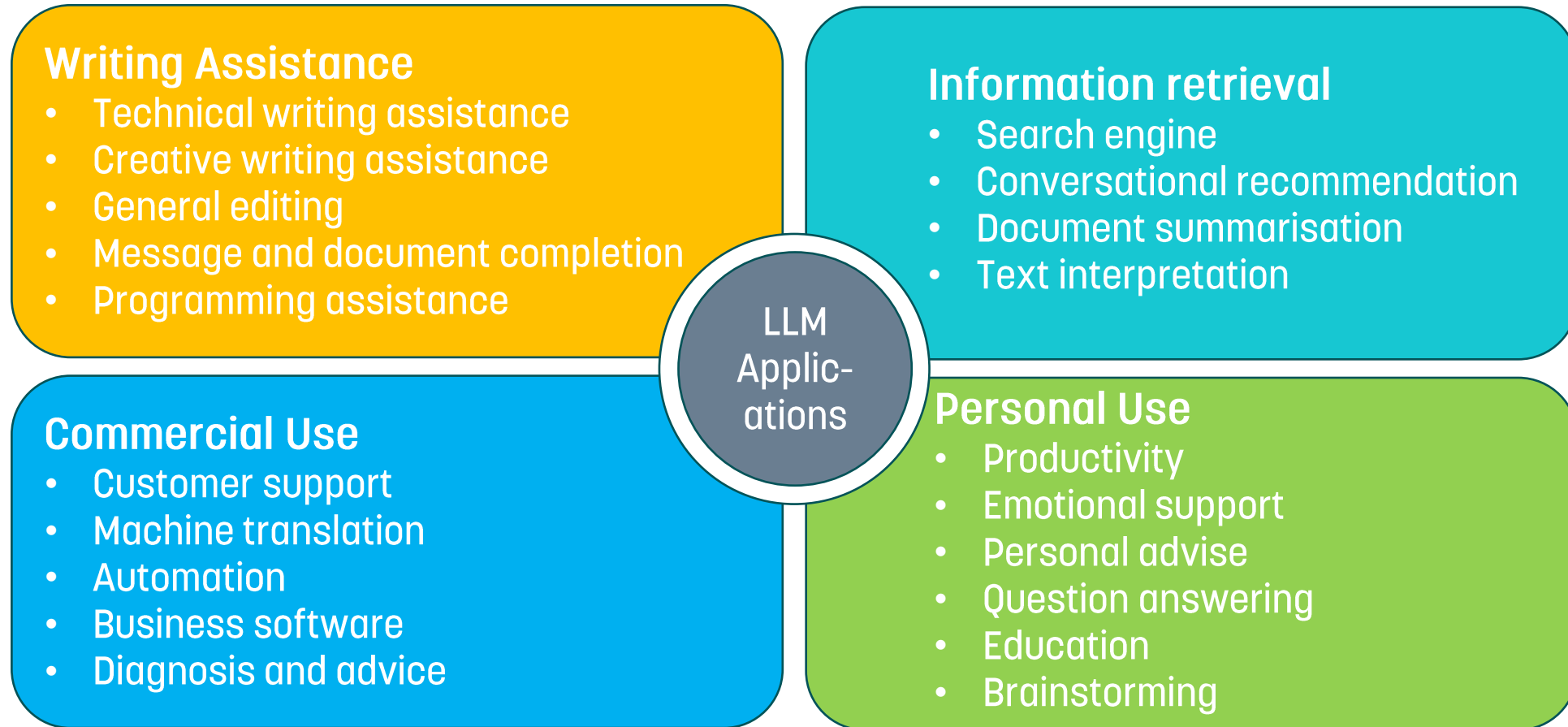
Zaharia et al. 2024. The Shift from Models to Compound AI Systems

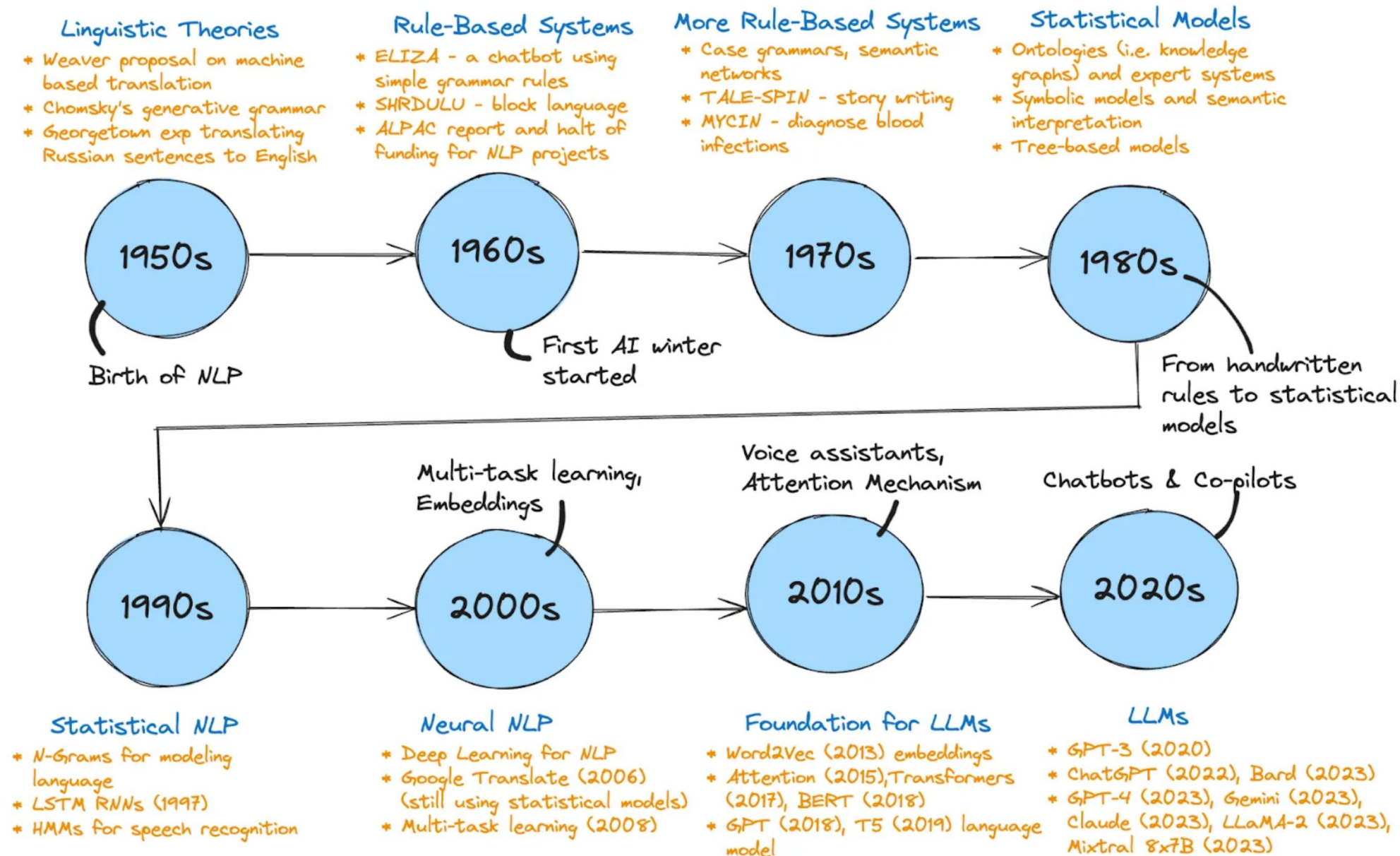
Generative AI

"Teddy bears mixing sparkling chemicals as mad scientists in a 'steampunk' style."



Large Language Model Applications





Language Modelling

- **Language modelling** is the task of predicting which word comes next in a sequence of words.
- More formally, given a sequence of words $w_1, \dots, w_t$ we want to know the probability of the next word, w_{t+1} :

$$P(w_{t+1} \mid w_1, \dots, w_t)$$

- We are assuming that w_{t+1} comes from a finite vocabulary V .

language models = classifiers

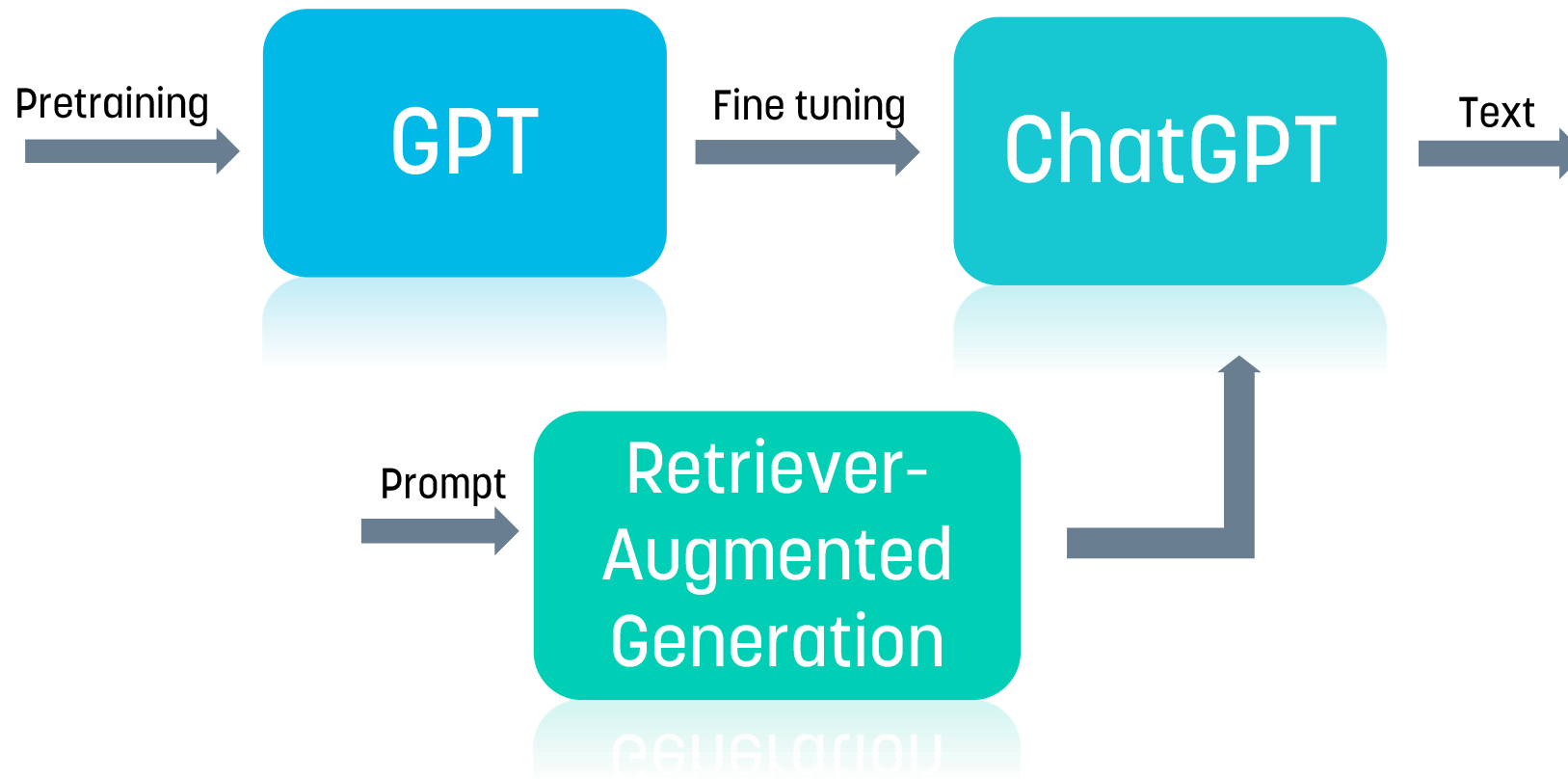
An Approximation

- A problem: computing $P(w \mid h)$ exactly is infeasible for arbitrary history h since language is **creative** and h might have never occurred before!
- Idea: the Markov assumption: approximate the history by just the last few words.

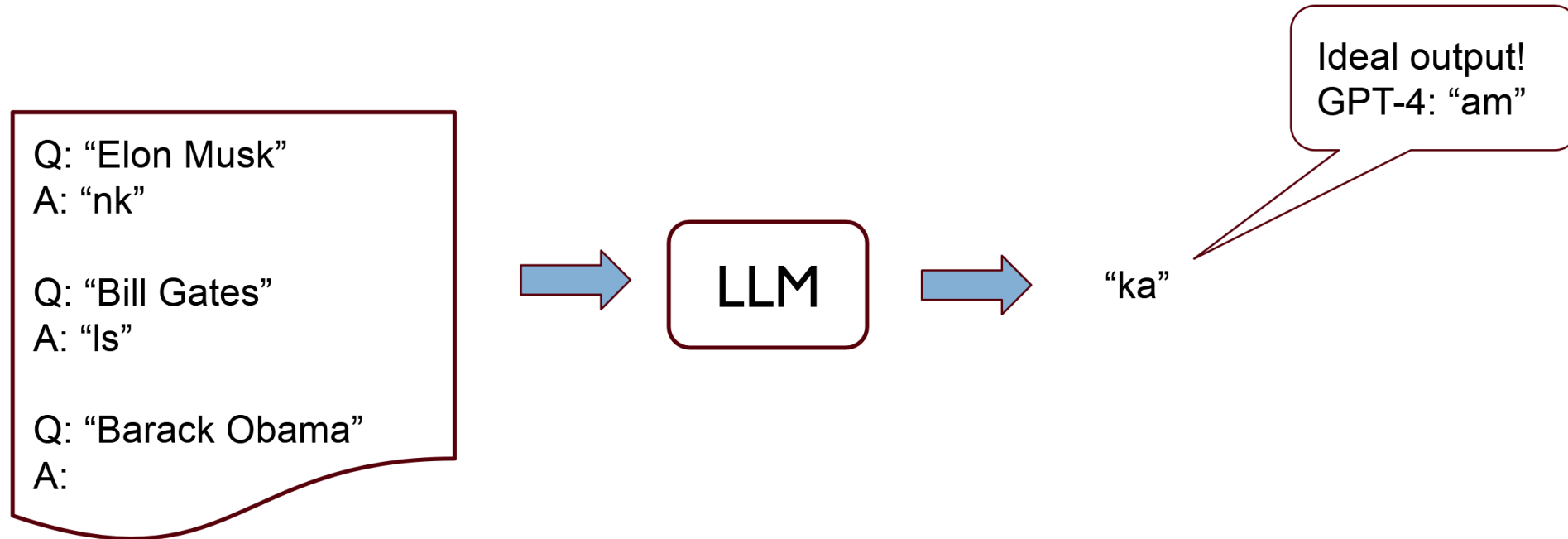
$$P(w_k|w_{1:k-1}) \approx P(w_k|w_{k-n+1:k-1})$$

- Example: n-gram models: look at n-1 words in the history. n=2 is bigrams, n=3 is trigrams, etc.
- An **n-gram** is a contiguous sequence of n words (or characters).
 - Sherlock **Holmes** had **sprung out** and seized the intruder **by the collar**.
unigram bigram trigram
- LLMs use **much** larger n , in the thousands or even millions!

How Does ChatGPT Work?



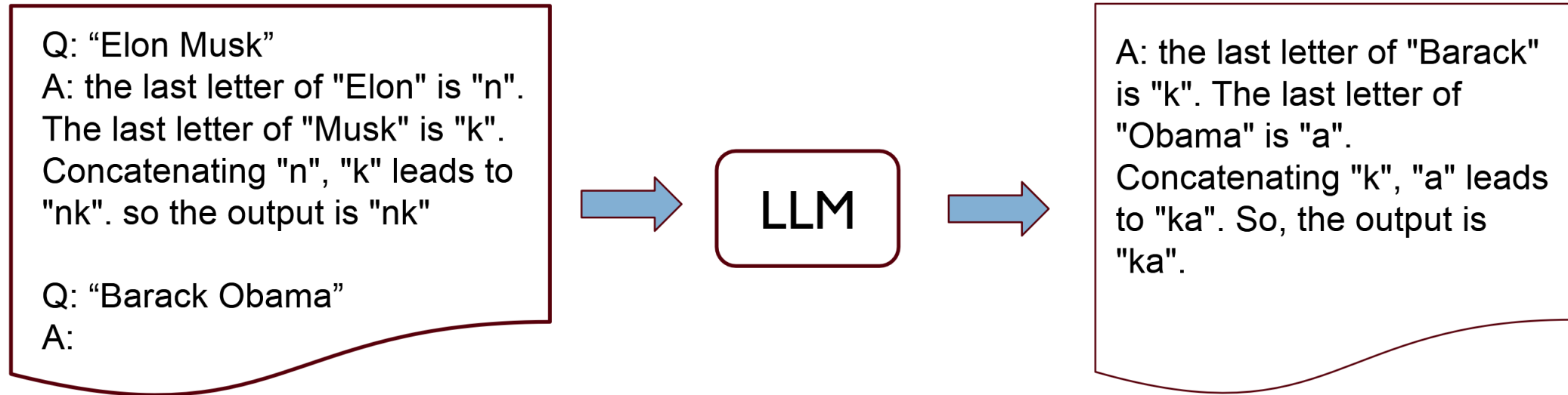
Few Shot Prompting



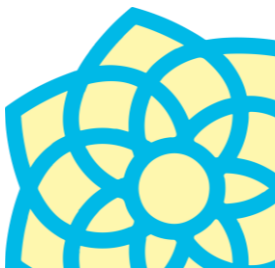
T. Brown et al. [Language Models are Few-Shot Learners](#). NeurIPS 2020.



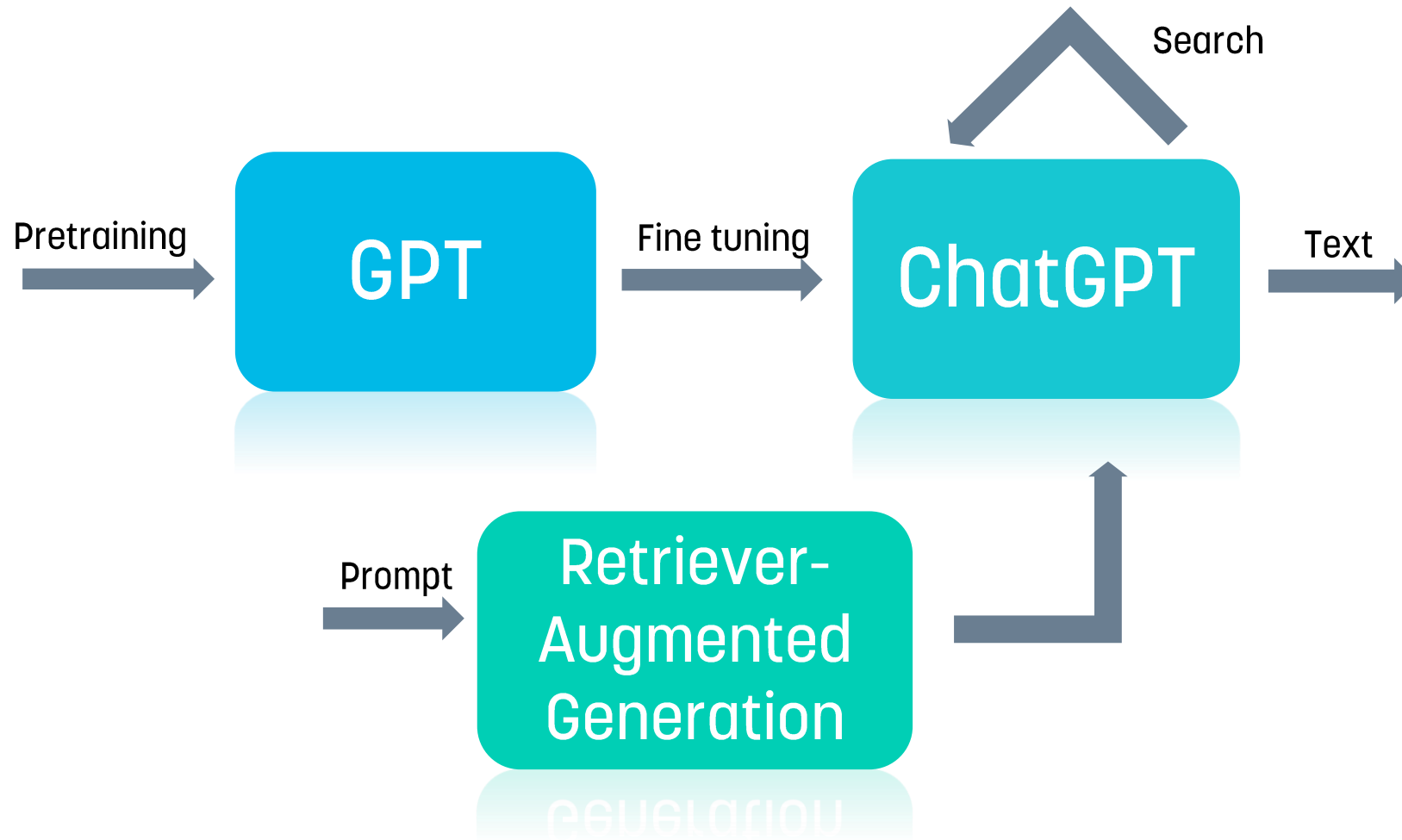
Chain of Thought Prompting



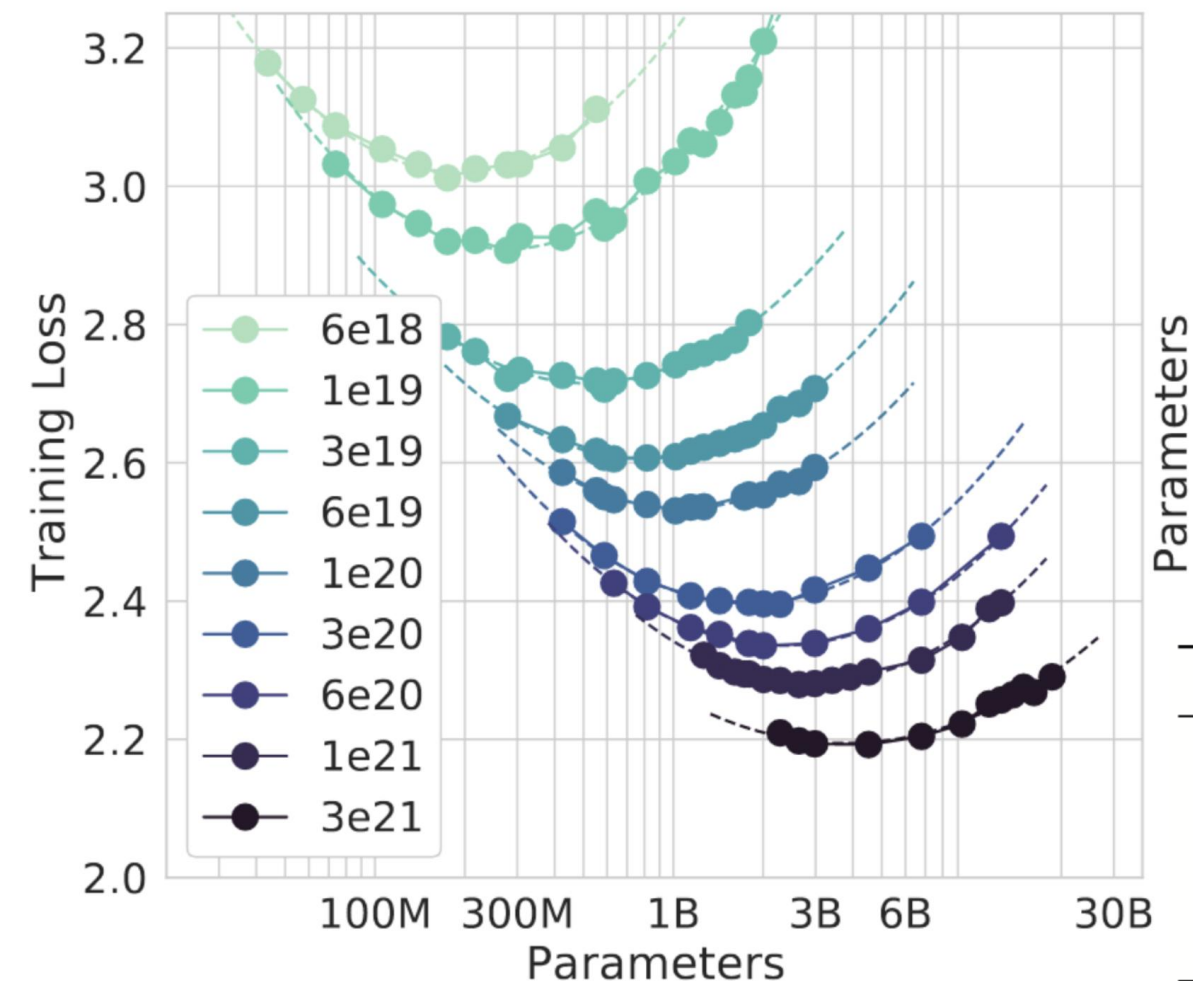
J. Wei et al. [Chain-of-Thought Prompting Elicits Reasoning in Large Language Models](#). NeurIPS 2022.



How Does ChatGPT-o1 Work?



Large Language Models Scaling Laws

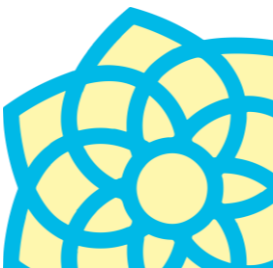
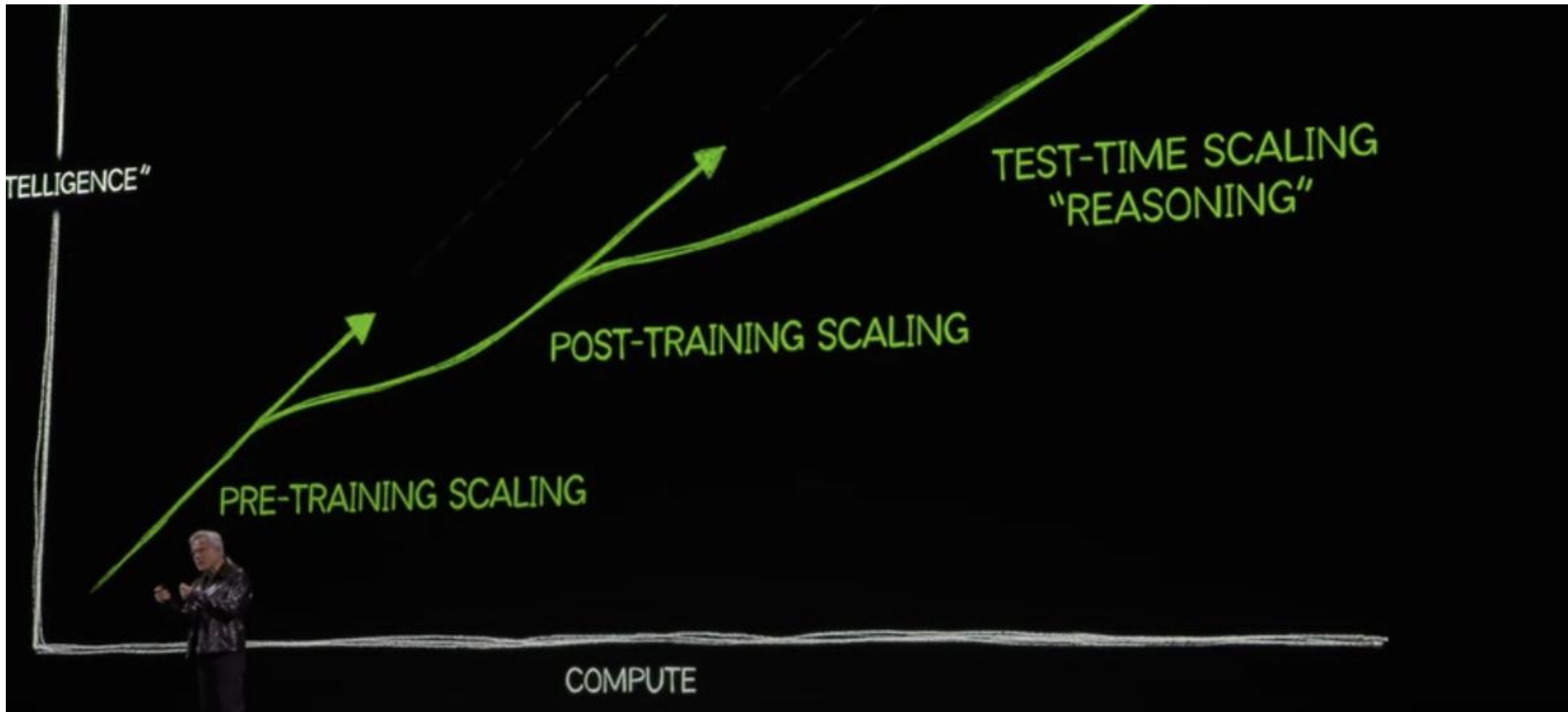


Parameters	FLOPs	FLOPs (in <i>Gopher</i> unit)	Tokens
400 Million	1.92e+19	1/29,968	8.0 Billion
1 Billion	1.21e+20	1/4,761	20.2 Billion
10 Billion	1.23e+22	1/46	205.1 Billion
67 Billion	5.76e+23	1	1.5 Trillion
175 Billion	3.85e+24	6.7	3.7 Trillion
280 Billion	9.90e+24	17.2	5.9 Trillion
520 Billion	3.43e+25	59.5	11.0 Trillion
1 Trillion	1.27e+26	221.3	21.2 Trillion
10 Trillion	1.30e+28	22515.9	216.2 Trillion

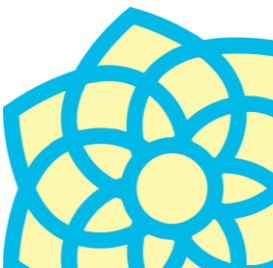
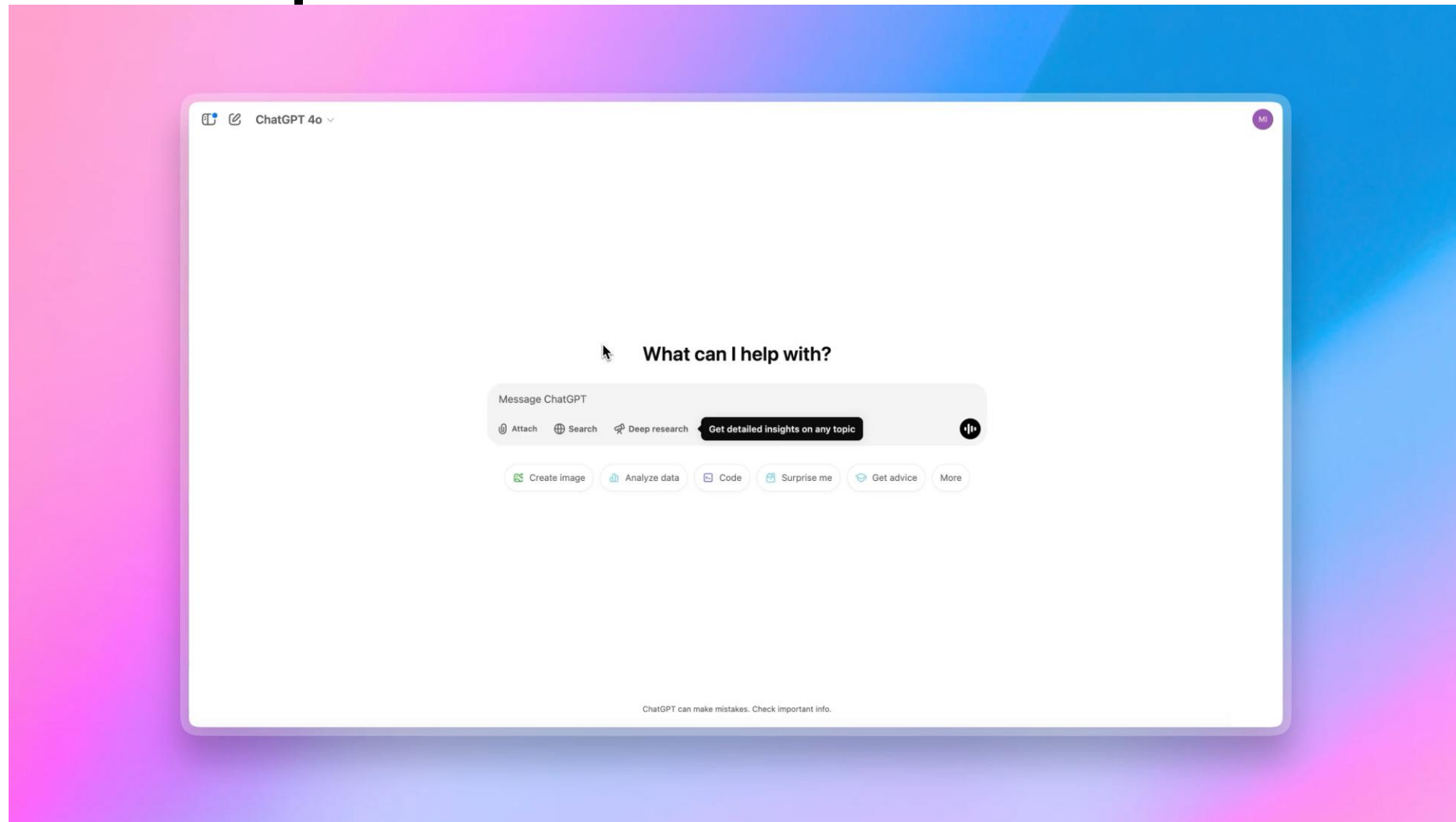
Model	Size (# Parameters)	Training Tokens
LaMDA (Thoppilan et al., 2022)	137 Billion	168 Billion
GPT-3 (Brown et al., 2020)	175 Billion	300 Billion
Jurassic (Lieber et al., 2021)	178 Billion	300 Billion
<i>Gopher</i> (Rae et al., 2021)	280 Billion	300 Billion
MT-NLG 530B (Smith et al., 2022)	530 Billion	270 Billion



Scaling – From Pre-training to Test-time



OpenAI Deep Research



From Generative to Agentic AI

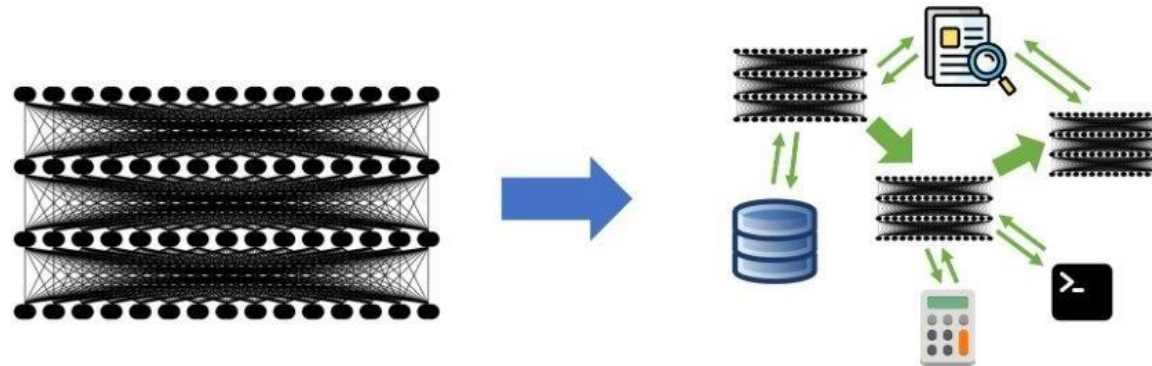
- Generative

- Generate content like text & image



- Agentic

- Execute complex tasks on behalf of human



Zaharia et al. 2024. The Shift from Models to Compound AI Systems

Examples of Agentic AI

- Personal assistants
- Autonomous robots
- Gaming agents
- Science agents
- Web agents
- Software agents



Creative Writing Coach

I'm excited to read your work and give you feedback to improve your skills.



Laundry Buddy

Ask me anything about st settings, sorting and ever laundry.

Game Time

I can quickly explain board games or card games to players of any skill level. Let the games begin!



Tech Advisor

From setting up a printer to troubleshooting a device, I'm here to help you step-by-step.



The Negotiator

I'll help you advocate for y get better outcomes. Bec negotiator.



Sticker Whiz

I'll help turn your wildest dreams into die-cut stickers, shipped to your door.



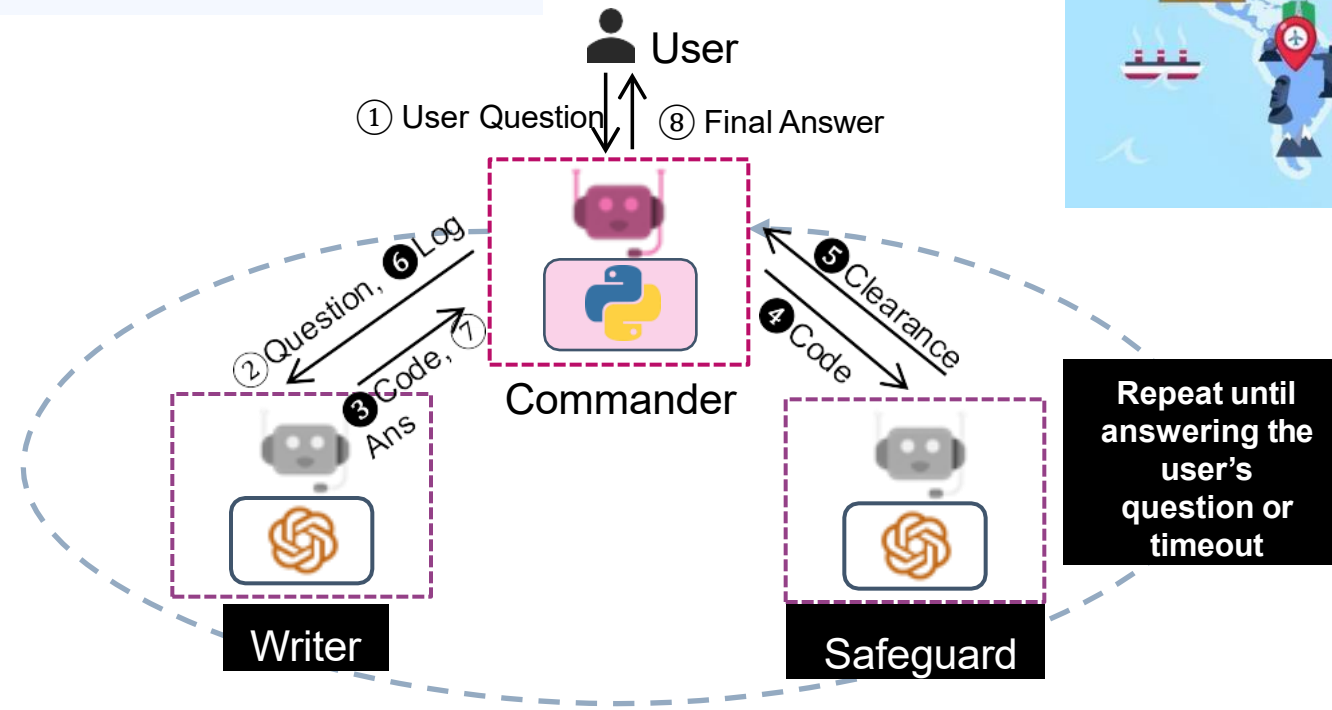
The Negotiator

I'll help you advocate for y get better outcomes. Bec negotiator.

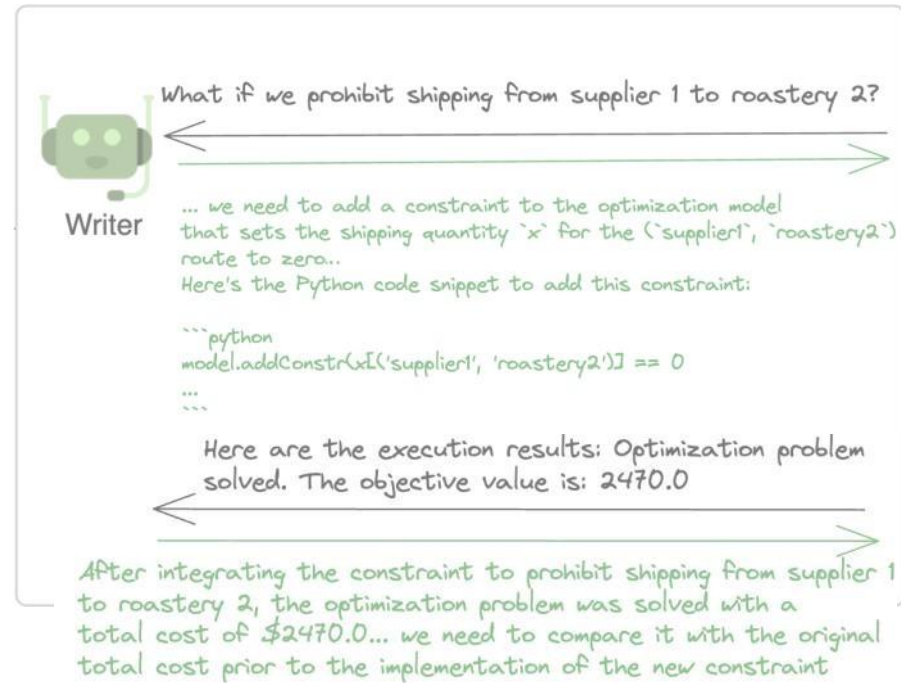
Key Benefits of Agentic AI

- Useful Interface
 - Natural interaction with human agency
- Strong Capability
 - Operate with minimal human intervention
- Useful Architecture
 - Intuitive programming paradigm

What if we prohibit shipping from supplier 1 to roastery 2?

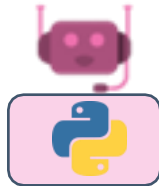


(Writer nested in Commander, Triggered by User)



User

What if we prohibit shipping from supplier 1 to roastery 2?



Commander

(Safeguard nested in Commander, Triggered by Writer)

Is the code safe?

```
python
model.addConstr(x['supplier1', 'roastery2']) == 0
...
```

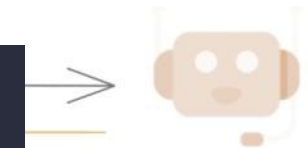
```
Optimize a model with 11 rows, 18 columns and 36 nonzeros
Model fingerprint: 0x8aa2c280
Variable types: 0 continuous, 18 integer (0 binary)
Coefficient statistics:
  Matrix range      [1e+00, 1e+00]
  Objective range   [2e+00, 1e+01]
  Bounds range      [0e+00, 0e+00]
  RHS range         [2e+01, 2e+02]
Found heuristic solution: objective 2900.0000000
Presolve time: 0.00s
Presolved: 11 rows, 18 columns, 36 nonzeros
Variable types: 0 continuous, 18 integer (0 binary)
Found heuristic solution: objective 2896.0000000

Root relaxation: objective 2.4700000e+03, 11 iterations, 0.00 seconds (0.00 work units)

  Nodes |   Current Node   |   Objective Bounds   |   Work
Expl Unexpl | Obj Depth IntInf | Incumbent   BestBd   Gap | It/Node Time
*    0     0           0  2470.0000000 2470.00000  0.00%   -    0s

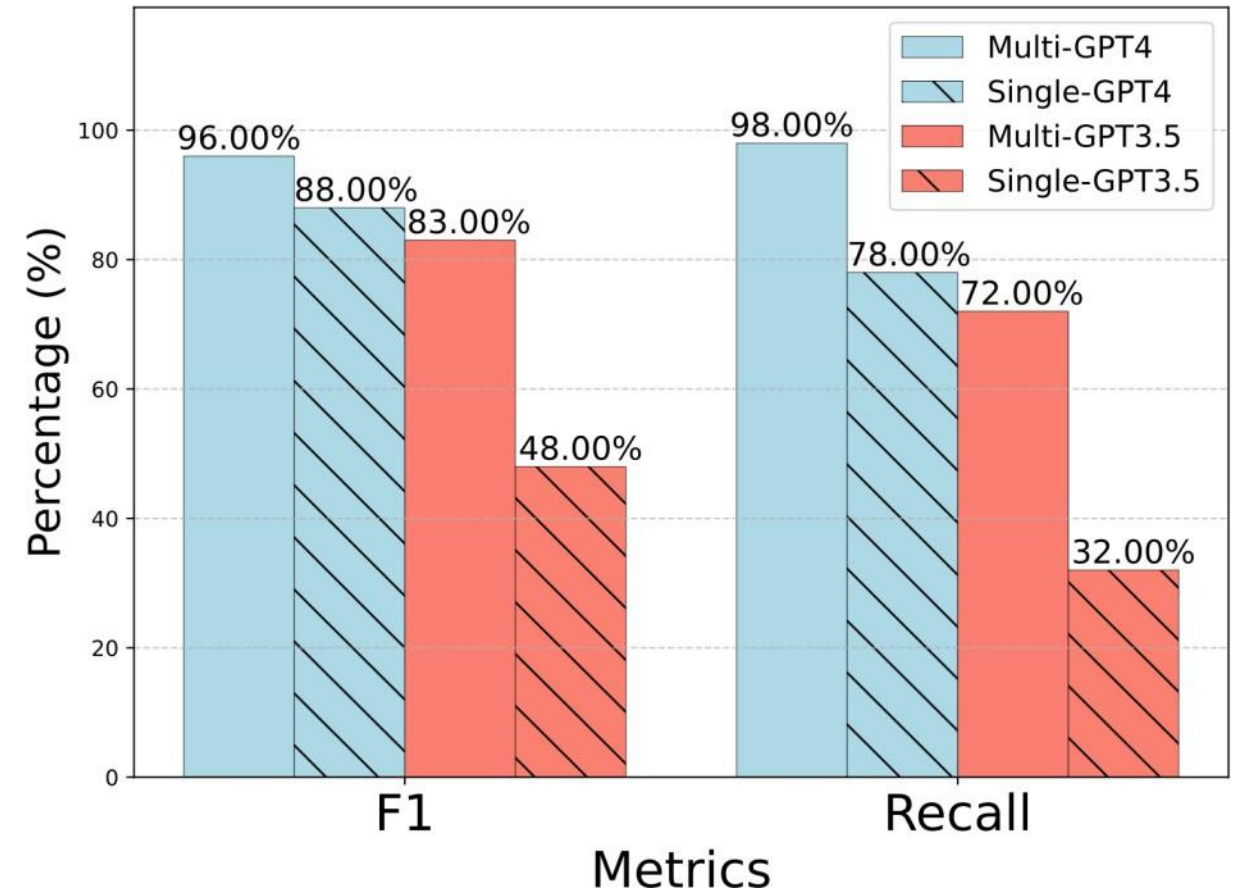
Explored 1 nodes (11 simplex iterations) in 0.00 seconds (0.00 work units)
Thread count was 14 (of 14 available processors)

Solution count 3: 2470 2896 2900
```



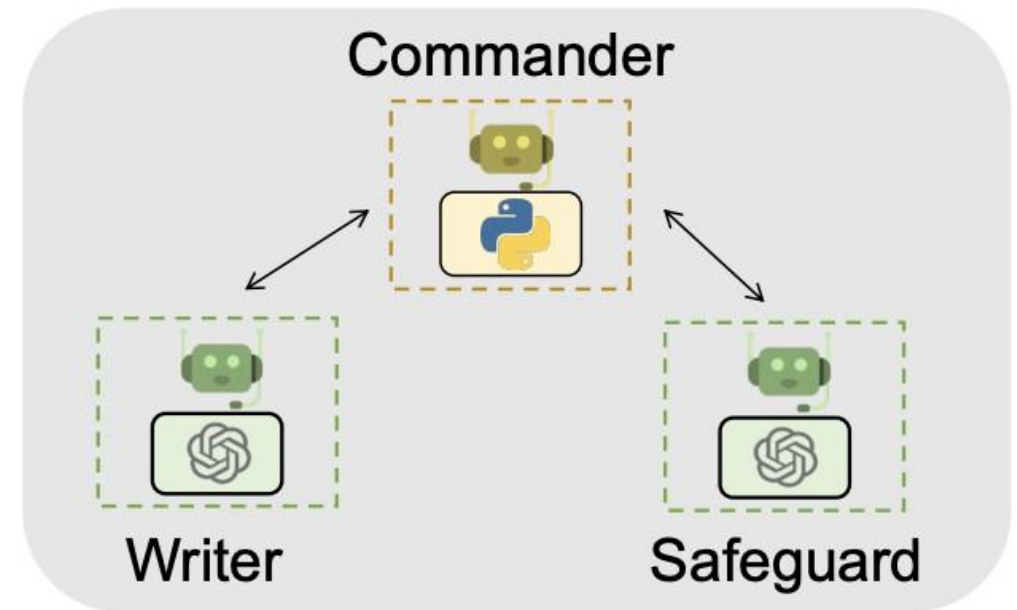
Agentic Programming

- Handle more complex tasks / Improve response quality
 - Improve over natural iteration
 - Divide & conquer
 - Grounding & validation



Agentic Programming

- Easy to understand, maintain, extend
 - Modular composition
 - Natural human participation
 - Fast & creative experimentation

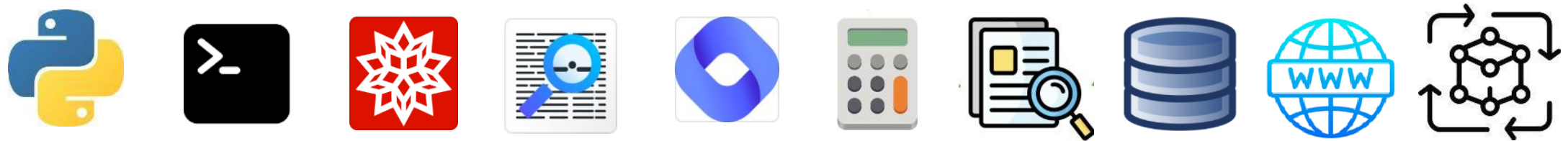
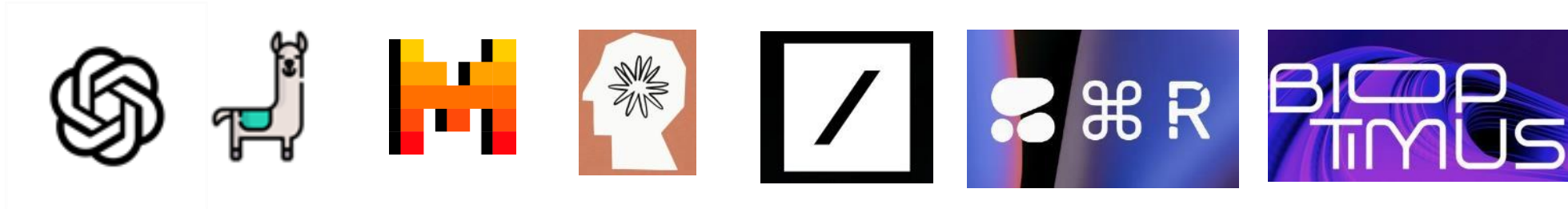


Agentic AI Framework Desiderata

- Intuitive unified agentic abstraction
- Flexible multi-agent orchestration
- Effective implementation of agentic design patterns
- Support diverse application needs

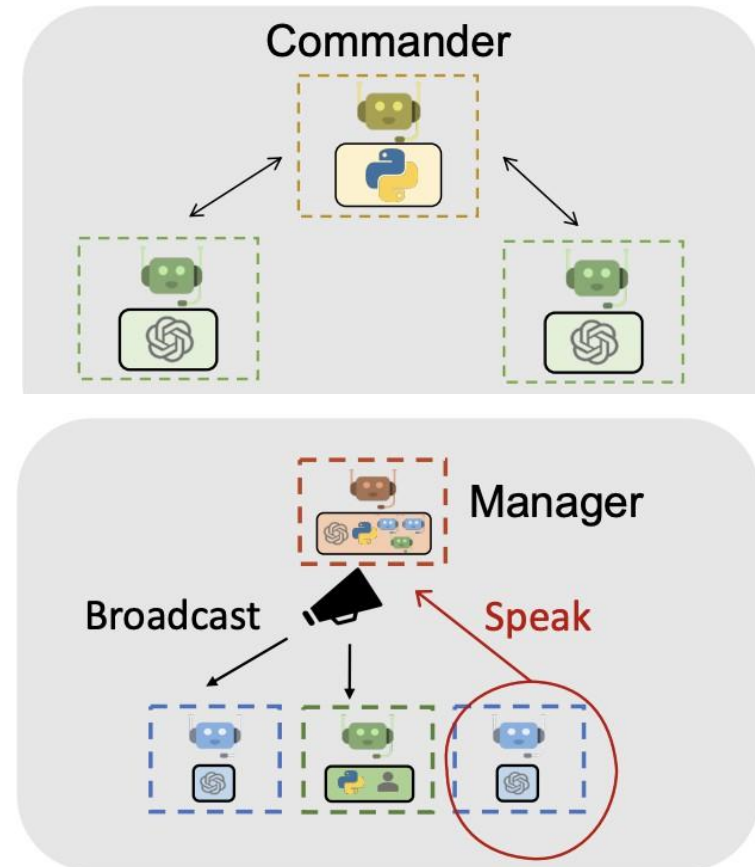
Agentic Abstraction

Unify models, tools, human for compound AI systems



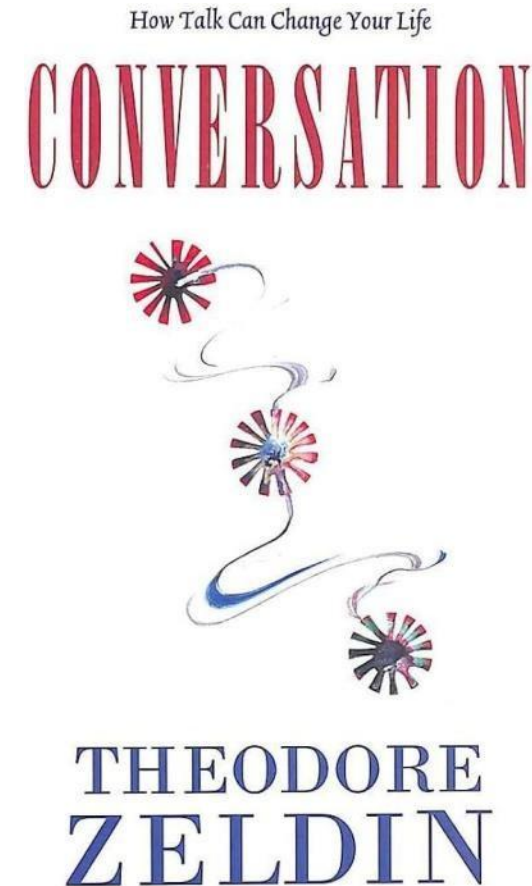
Multi-Agent Orchestration

- Static/dynamic
- NL/PL
- Context sharing/isolation
- Cooperation/competition
- Centralized/decentralized
- Intervention/automation



Agentic Design Patterns

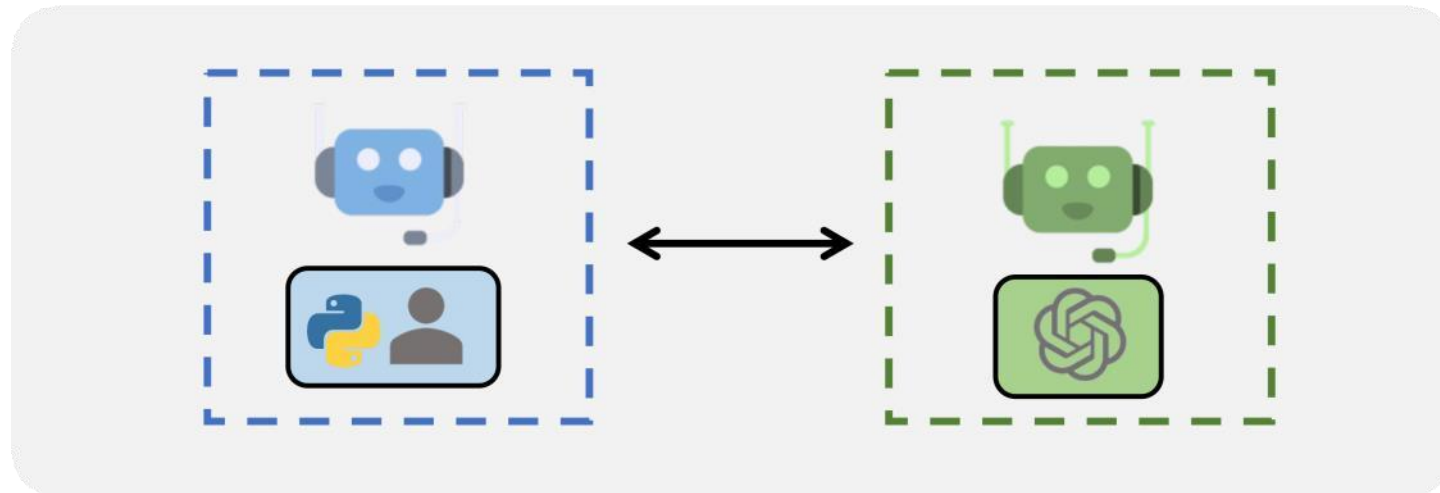
- Conversation
- Prompting & reasoning
- Tool use
- Planning
- Integrating multiple models, modalities and memories



Examples of Agentic AI Frameworks

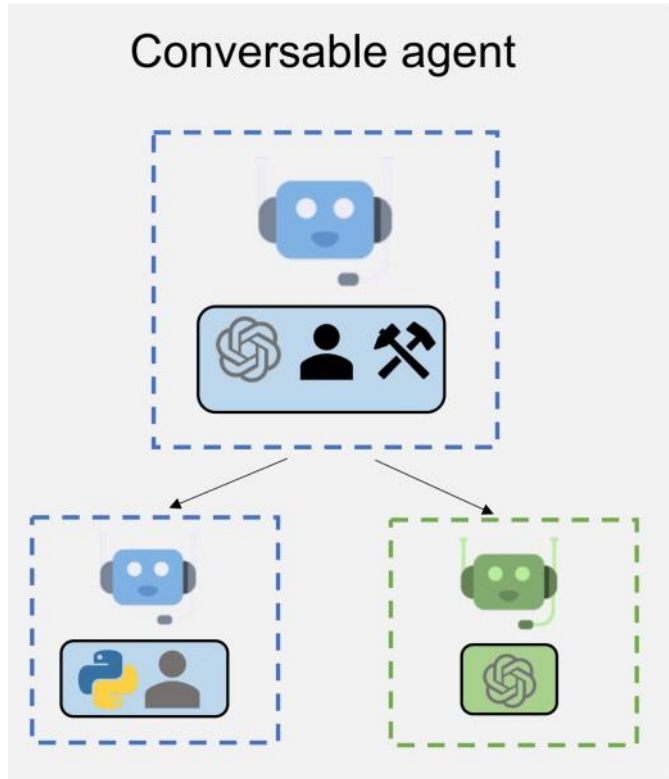
- AutoGen
 - Multi-agent conversation programming
 - Comprehensive & flexible
 - Integrable with other frameworks like OpenAI Assistant, LlamaIndex, LangChain
- Langchain-based
 - Langgraph
 - Graph-based control flow
 - CrewAI
 - High-level static agent- task workflow

AutoGen: A programming framework for agentic AI



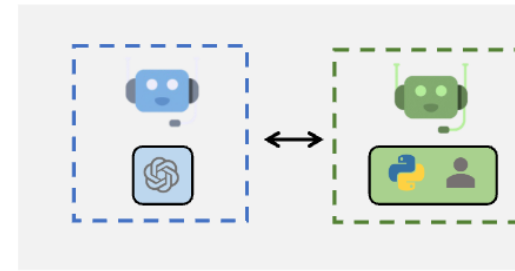
Initially developed in FLAML (Nov 2022)
Spined off to a standalone repo (October 2023)
Standalone GitHub organization AutoGen-AI (August 2024)

Define agents:
Conversable & Customizable

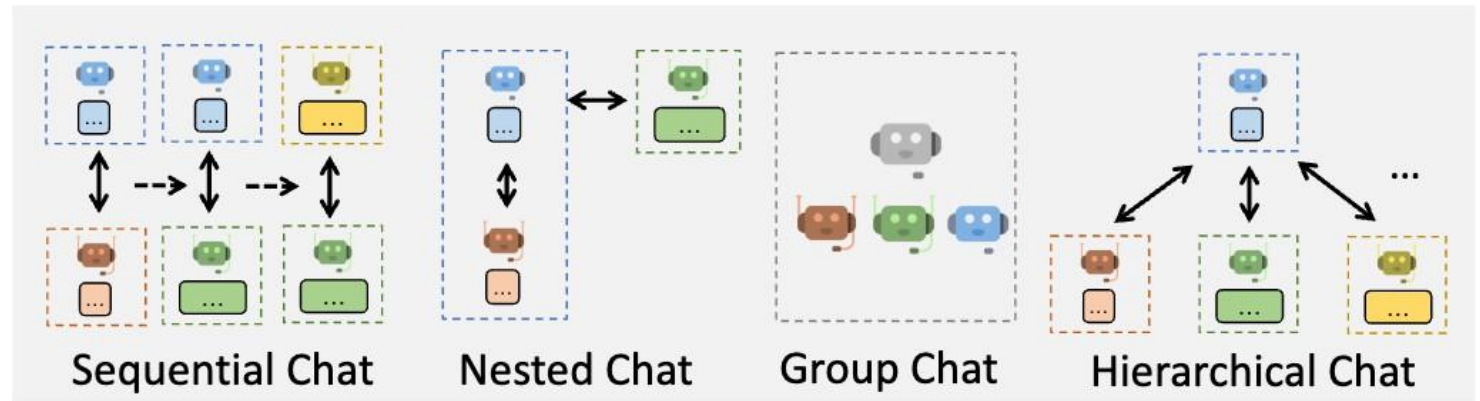


Agent Customization

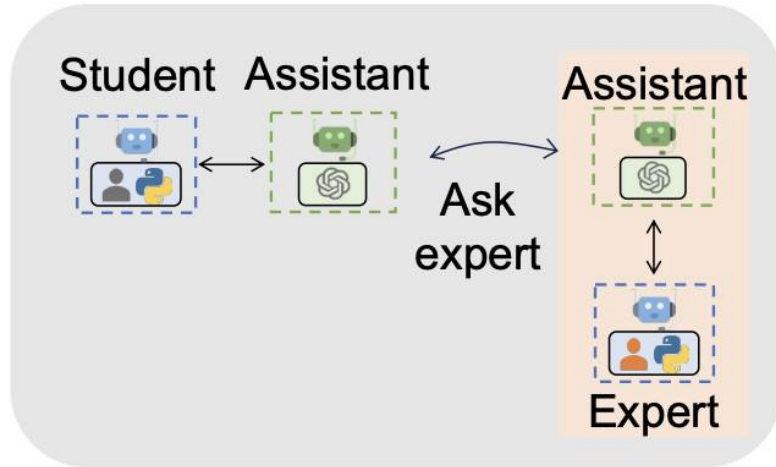
Get them to talk:
Conversation Programming



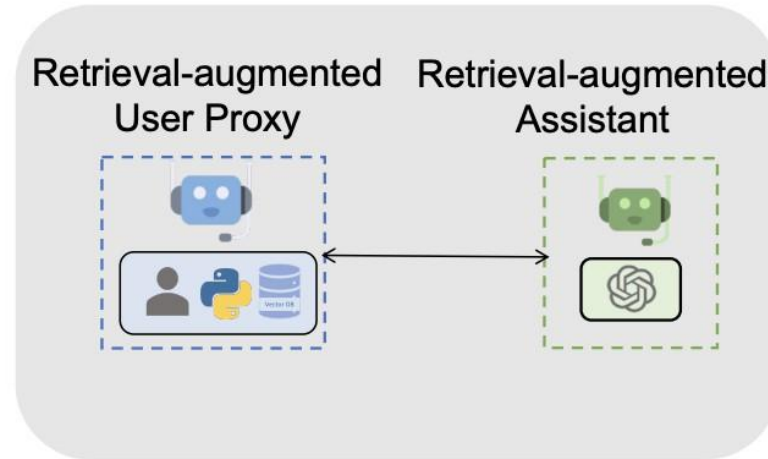
Multi-Agent Conversations



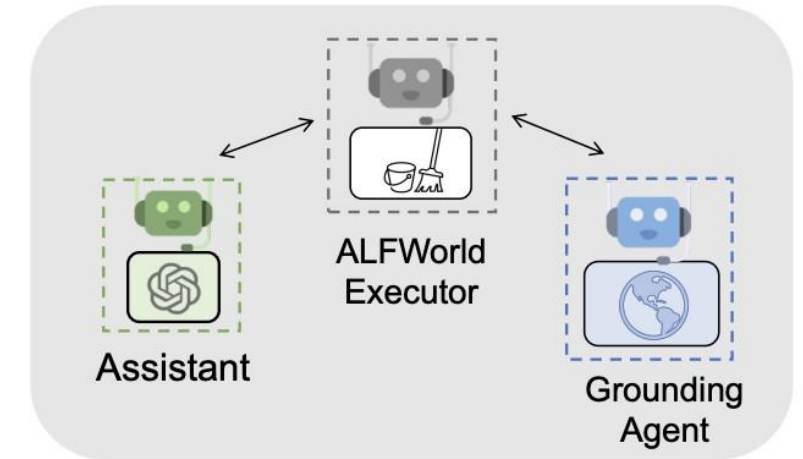
Flexible Conversation Patterns



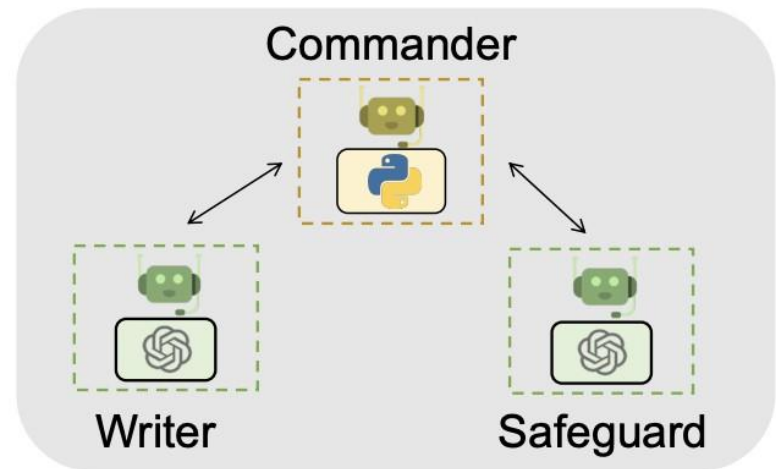
A1. Math Problem Solving



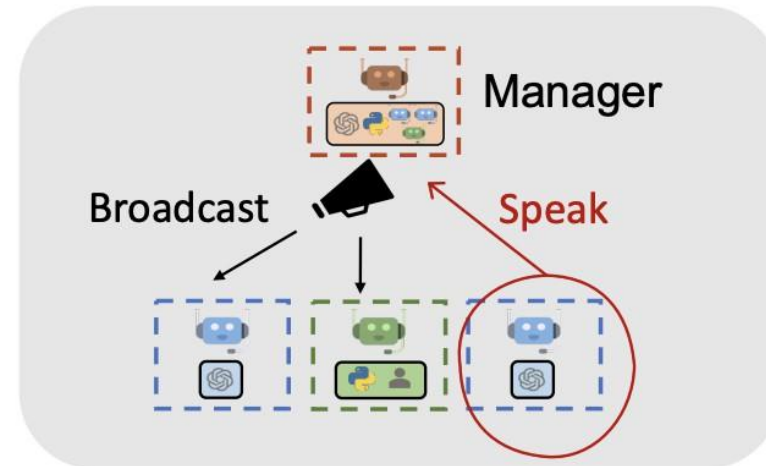
A2. Retrieval-augmented Q&A



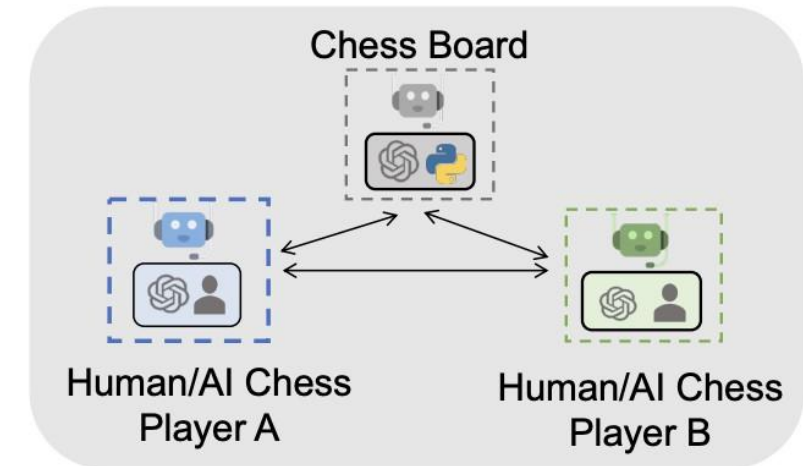
A3. Decision Making in Embodied Agents



A4. Supply-Chain Optimization



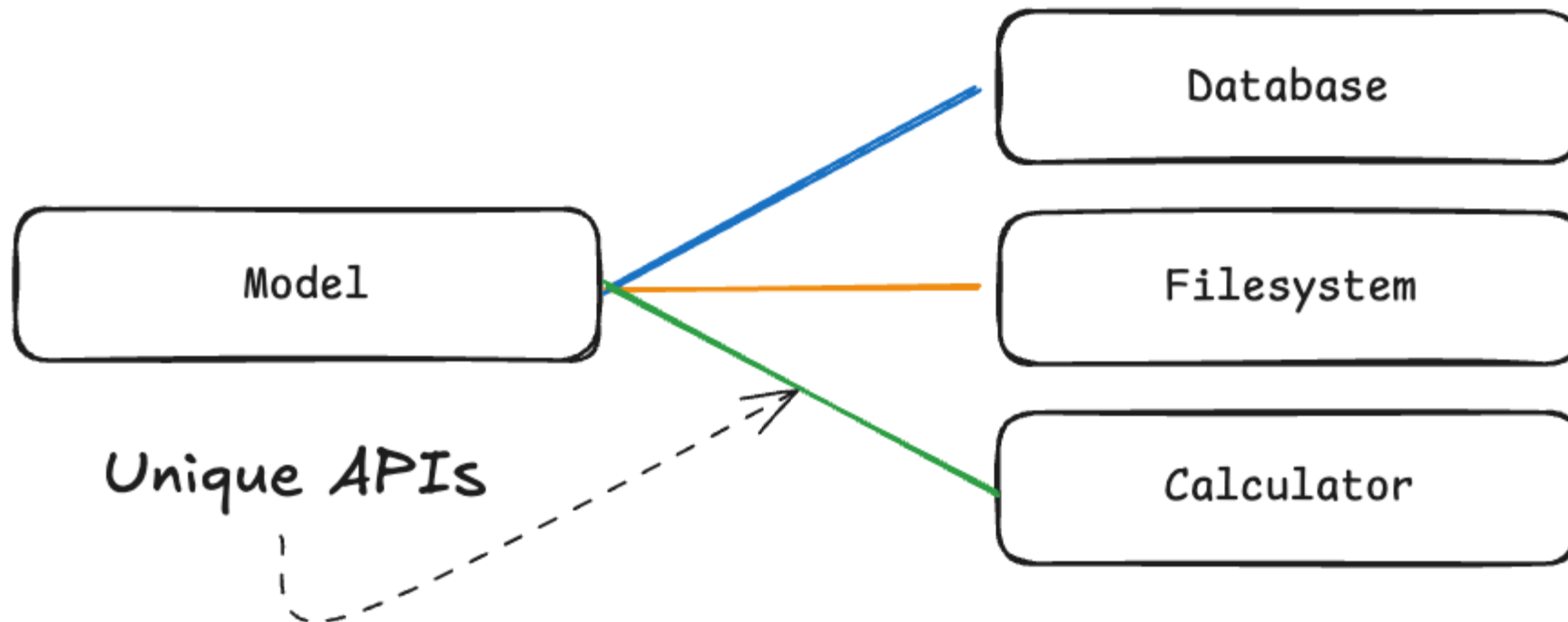
A5. Dynamic Task Solving with Group Chat



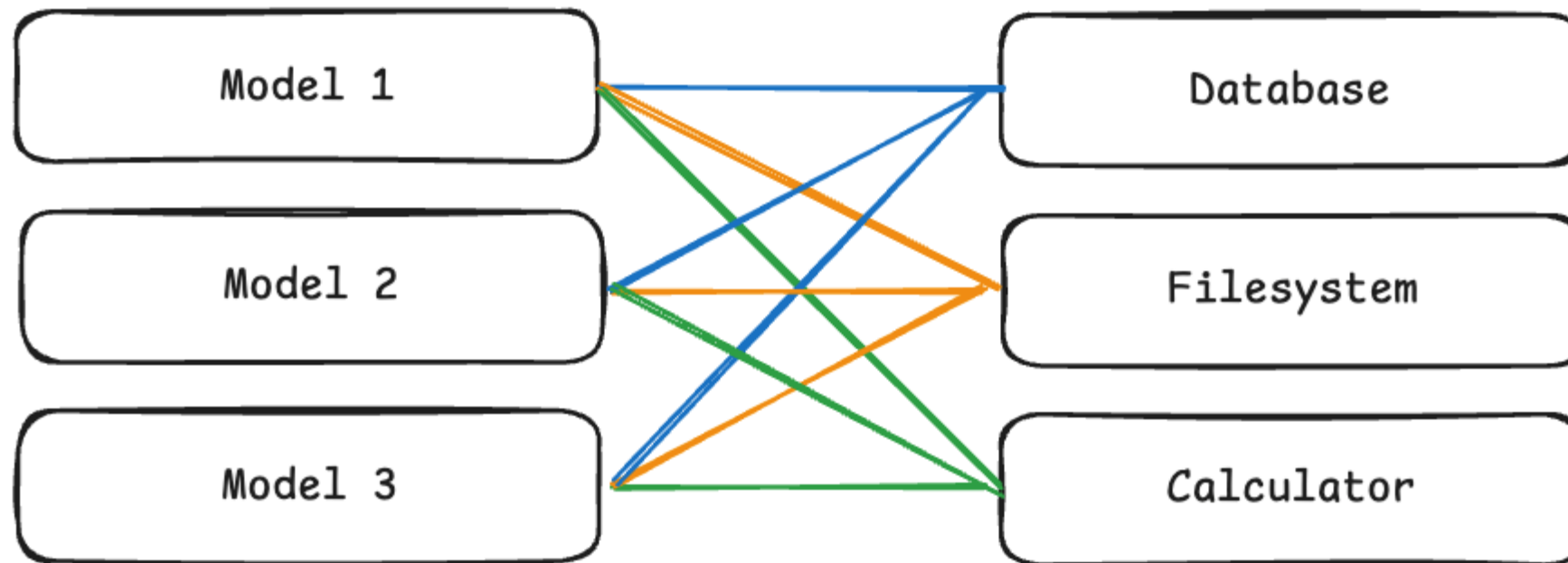
A6. Conversational Chess

For more examples: <https://autogen-ai.github.io/autogen/docs/notebooks>

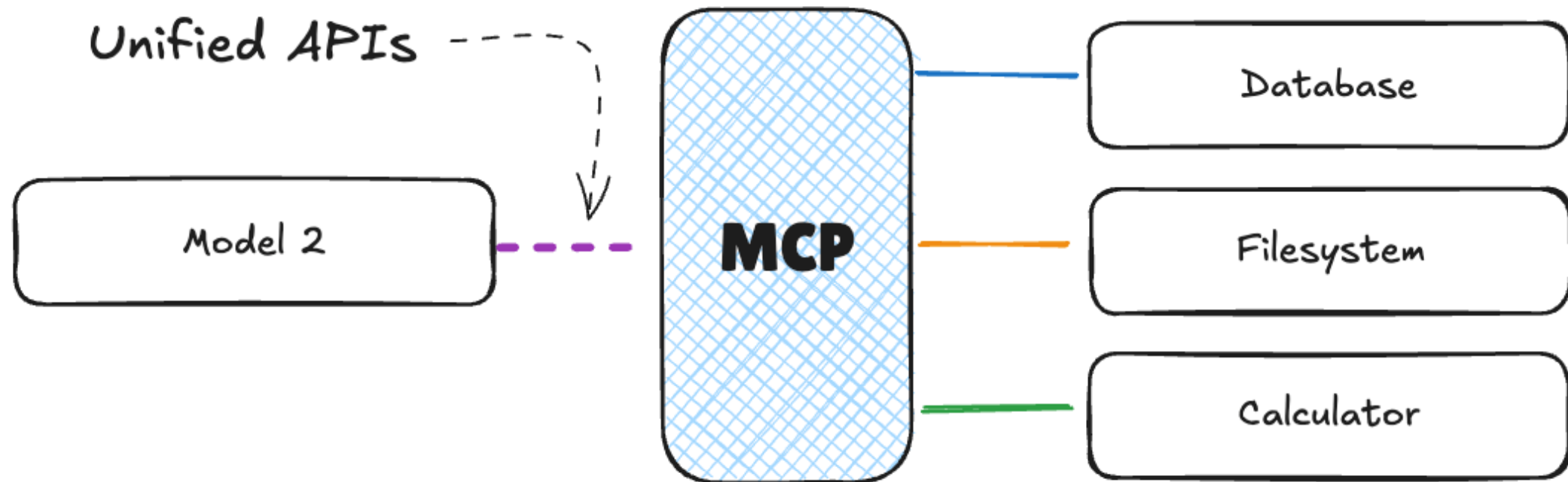
Model Context Protocol (MCP)



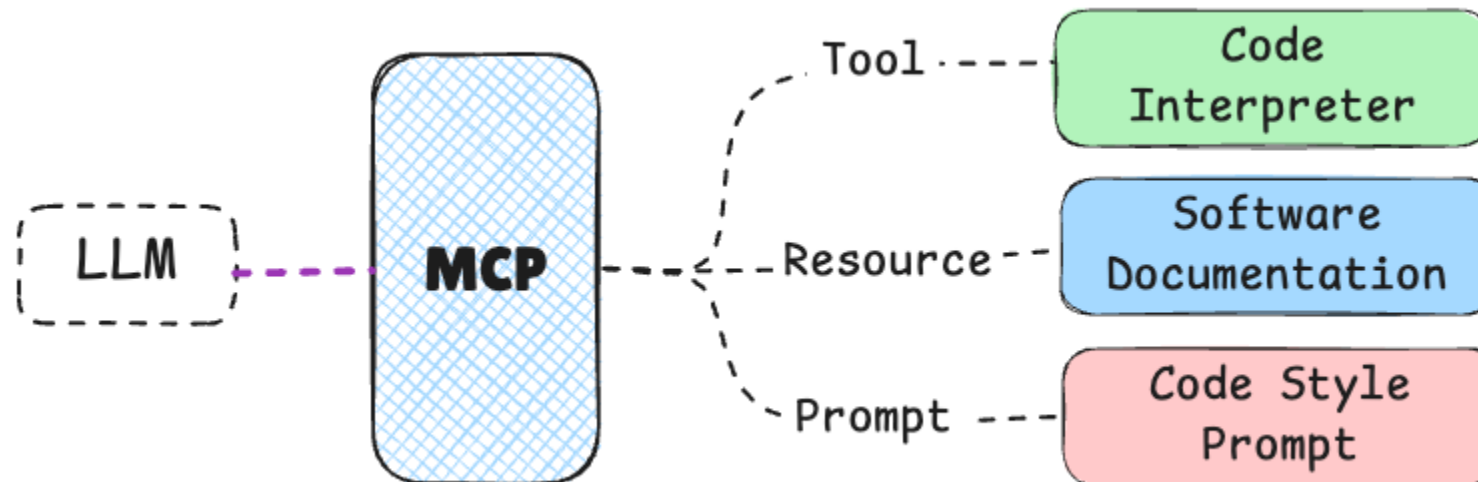
Model Context Protocol (MCP)



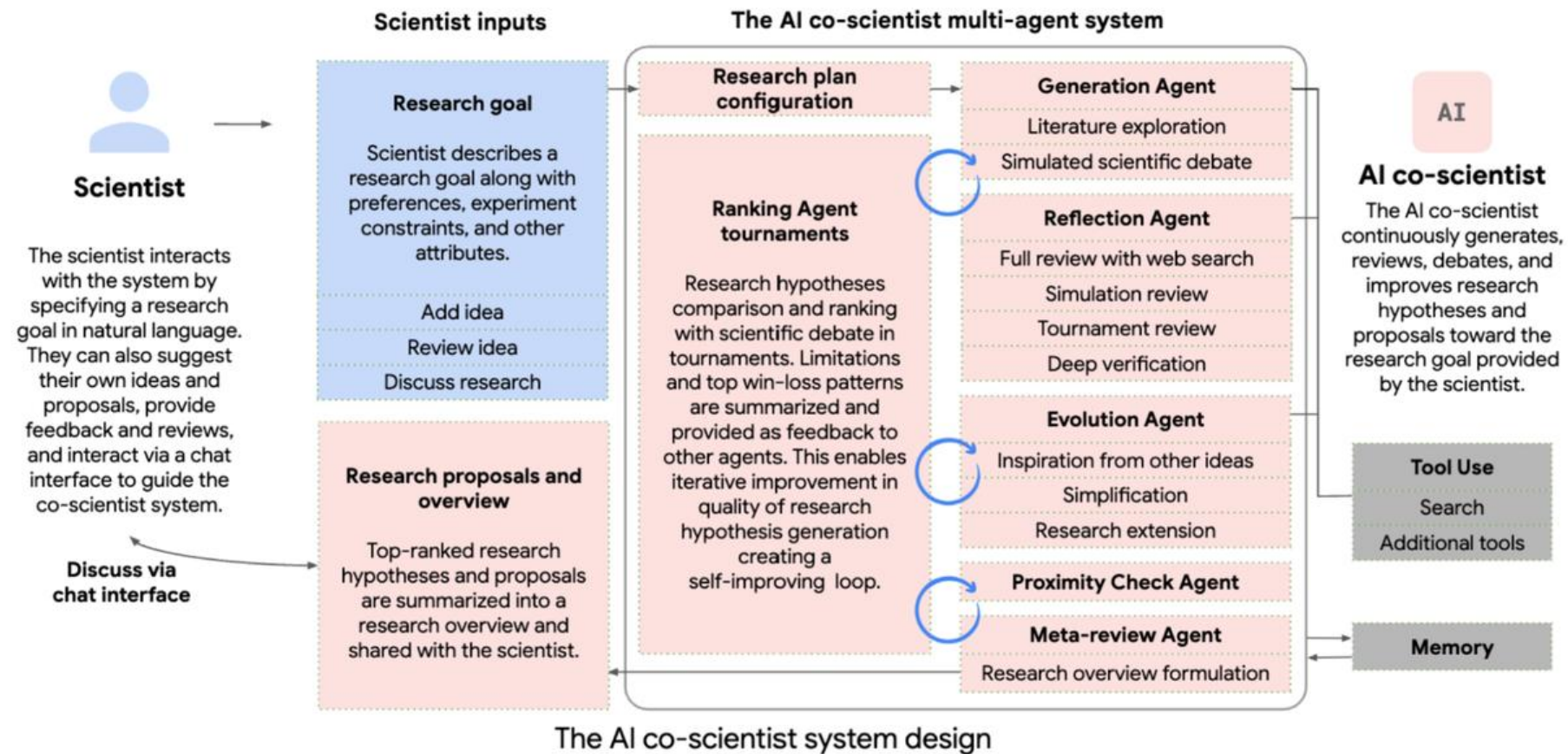
Model Context Protocol (MCP)



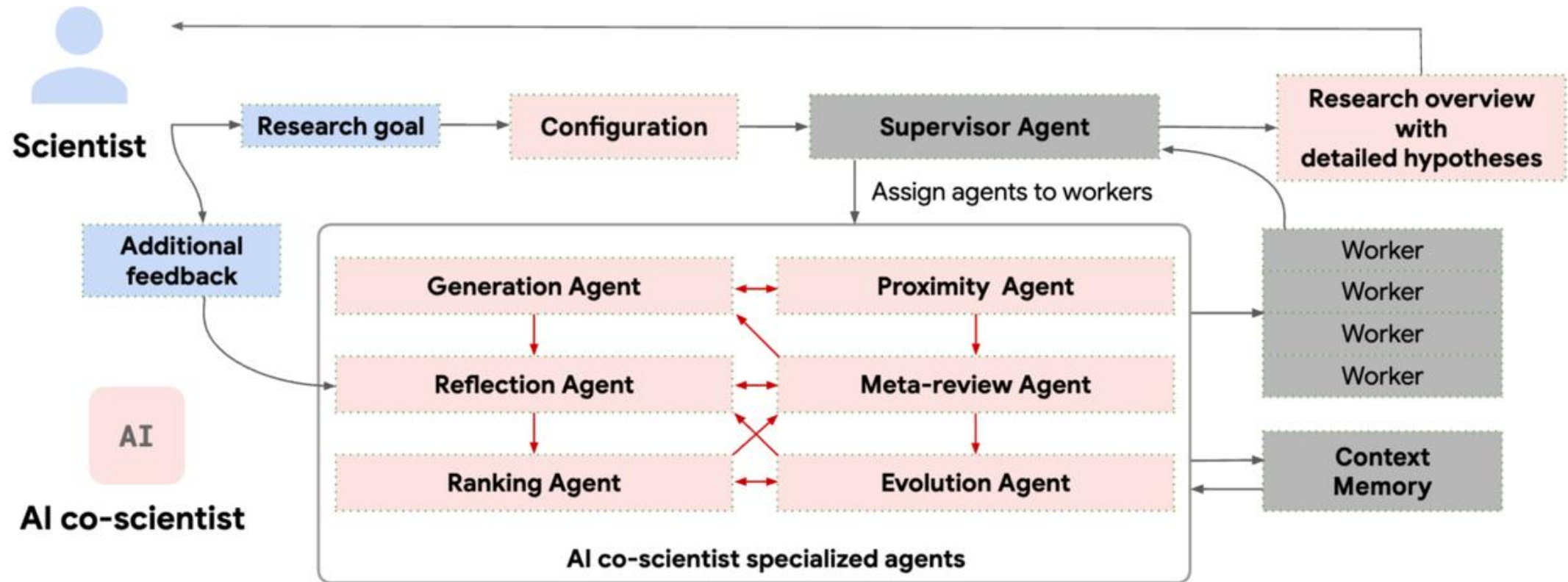
Model Context Protocol (MCP)



Google AI Co-Scientist



Google AI Co-Scientist



**TDDE13 MAS LE3 HT2025:
Large Language Models
Agentic AI
Model Context Protocol**

www.ida.liu.se/~TDDE13